1-2021

Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years.

Gabrielle Strouse
University of South Dakota, gabrielle.strouse@usd.edu

Jennifer Samson
Queens University - Charlotte

Recommended Citation
Strouse, G. A., & Samson, J. (2021). Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years. Child Development, 92(1), e20-e38. https://doi.org/10.1111/cdev.13429

This is the accepted version of the following article:

Strouse, G. A., & Samson, J. E. (2021). Learning from video: A meta-analysis of the video deficit in children ages 0 to 6 years. *Child Development, 92(1), e20-e38. https://doi.org/10.1111/cdev.13429*

Learning From Video: A Meta-Analysis of the Video Deficit in Children Ages 0 to 6 Years

Gabrielle A. Strouse, University of South Dakota

Jennifer E. Samson, Queens University of Charlotte

Author Note

Gabrielle A. Strouse, Division of Counseling and Psychology in Education, University of South Dakota; and Center for Brain and Behavior Research, University of South Dakota; Jennifer E. Samson, Department of Psychology, Queens University of Charlotte.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Thank you to Emily E. Clark and Robyn Warner Lachman for their work on reliability coding, as well as Emily Tanner-Smith for guidance with analyses, and the reviewers for feedback on earlier versions.

Correspondence concerning this article should be addressed to Gabrielle Strouse, University of South Dakota School of Education, 414 E. Clark St., Vermillion, SD 57069. Phone: 605-667-5846 Email: gabrielle.strouse@usd.edu

Abstract

Young children often learn less from video than face-to-face presentations. Meta-regression models were used to examine the average size of this difference (video deficit) and investigate moderators. An average deficit of about half of a standard deviation was reported across 122 independent effect sizes from 59 reports, involving children ages 0-6 years. Moderator analyses suggested 1) the deficit decreased with age, 2) object retrieval studies showed larger deficits than other domains, and 3) there was no difference between studies using live versus prerecorded video. Results are consistent with a multiple-mechanism explanation for the deficit. However, the analyses highlighted potential quality and publication bias issues that may have resulted in overestimation of the effect and should be addressed by future researchers.

Keywords: video deficit, meta-analysis, early childhood

Learning from Video: A Meta-Analysis of the Video Deficit in Children Ages 0 to 6 Years

Young children often demonstrate less learning after viewing a video presentation than after viewing a face-to-face presentation of the same information (Anderson & Pempek, 2005; Barr, 2010). This relative difference in children's learning is referred to as the *video deficit* (Anderson & Pempek, 2005). Several reviews of the topic illustrate the stability of the deficit across different content domains and learning outcomes, especially when children are in their second year of life (Barr, 2010; Kirkorian, 2018; Krcmar, 2010; Troseth, 2010). However, researchers have provided inconsistent estimates regarding the breadth of ages and contexts in which the effect consistently appears. Understanding these moderators is important not just for researchers attempting to determine the cause of the deficit, but also to policy makers, as the video deficit is one of multiple factors referred to in policy statements encouraging parents to limit or avoid exposing their young children to screen-based media (Canadian Paediatric Society, 2017; Council on Communications and Media, 2016). The purpose of the current study is to quantify the average size of the video deficit for children ages 0 to 6 years and explore age, use of live versus pre-recorded video, and learning domain as moderators.

Causes of the Video Deficit

Researchers have proposed several overlapping causes for the video deficit, focusing on differences between learning in video and face-to-face contexts. Differences include the perceptual features and social cues available in each context, the need for symbolic thinking or transfer, and prior experience with face-to-face and video interactions. These differences in learning contexts result in differential memory demands and conceptual challenges to learning.

Perception and memory. Video is two-dimensional, which results in significant perceptual differences between video and face-to-face learning contexts. The ability to use depth cues such as motion parallax, texture and shadow gradients, and stereopsis is limited from video presentations (Schmitt & Anderson, 2002). Because of these limitations, memories of video presentations that children encode may be less detailed than memories from face-to-face presentations, and encoding may take more time and cognitive resources (Carver, Meltzoff, & Dawson, 2006; Kirkorian et al., 2016; Schmitt & Anderson, 2002). Less detailed memories encoded from video may result in fewer retrieval cues available for children to access during later testing (Barr, 2010). In addition, video memories may mismatch future in vivo situations, making transfer more difficult (Barr, 2010). Overcoming these challenges may be cognitively demanding and require more working memory than using information learned face to face (Barr et al., 2016; Choi, Kirkorian, & Pempek, 2018; Kirkorian, 2018).

Social information. Video also lacks social cues that are common in learning situations. For example, infants are more likely to learn new words from parents who provide them with responses that are temporally and conceptually based on infants' actions, didactic (informative and referential), embodied (coordinated physically and verbally), and targeted to infants' current knowledge and developmental level (Tamis-LeMonda, Kuchirko, & Song, 2014). Pre-recorded videos, however, are not responsive to children's actions or scaffolded to be individually developmentally appropriate. Referential cues that are present, like an on-screen speaker's gaze, may be more difficult for children to follow. As a result, videos may not be as informative as face-to-face interactions because they are relatively socially impoverished (Krcmar 2010; Kuhl, 2007). In addition, children who are typically used to learning in socially responsive face-to-face situations may not understand the communicative intent of speakers on video (Kuhl, 2007). Contingency between a teacher and learner, such as eye contact and taking turns, signals learners that they are being taught, and that information is relevant and generalizable (Gergely, Egyed, Király, 2007). Without these cues, children may learn to conceptualize video as "not real" (Troseth, 2010, p. 164) or relevant, and fail to devote attention and cognitive resources to learning video information (Strouse, Troseth, O'Doherty, & Saylor, 2018).

Symbolic thinking. Finally, using information learned from video may also be challenging because a mature understanding of video relies on symbolic thinking (Troseth, 2010; Troseth, Flores, & Stuckelman, 2019). Tasks in which children are asked to generalize or transfer information from video to new situations require them to recognize the relation between the video event and the real event. Dual representation, or the ability to mentally represent objects or events on video as images on screen *as well as* representations of real-world objects and events, can support children in using information from video in real-world contexts (Troseth, 2010; Troseth & DeLoache, 1998). Experiences during which children can more easily link video experiences with real events (e.g., seeing themselves on live video, or having an adult explicitly point out the relation) and practice transferring from video to the real world may support their awareness of the video-reality connection and their usage of information learned from video in real-world contexts (Strouse & Troseth, 2014; Troseth, Casey, Lawver, Walker, & Cole, 2007; Troseth & DeLoache, 1998; Troseth et al., 2019).

Empirical evidence provides support for each of these mechanism's role in the deficit, but also suggests that none completely explains the deficit alone. For example, a study in which toddlers performed differently when they knowingly watched a video than when they were tricked into believing they were watching an in vivo event through a window (but were really watching a video) provides evidence that the deficit is not purely due to video's perceptual mismatch with in vivo events (Troseth & DeLoache, 1998). However, when testing and learning situations are perceptually matched, transfer is much higher than when they are mismatched

(Moser et al., 2015; Zack, Barr, Gerhardstein, Dickerson, & Meltzoff, 2009; Zack, Gerhardstein, Meltzoff, & Barr, 2013). Similarly, adding social cues to video through the use of live video feeds has supported learning in some studies (Myers, LeWitt, Gallo, & Maselli, 2017; Nielsen, Simcock, Jenkins, 2008; Roseberry, Hirsh-Pasek, & Golinkoff, 2014; Troseth, Saylor, & Archer, 2006) but not others (Myers, Crawford, Murphy, Aka-Ezoua, & Felix, 2018; Troseth, Strouse, Verdine, & Saylor, 2018; Strouse et al., 2018). In addition, conditions designed to highlight the relation between video and reality have supported learning in some paradigms but not all (Strouse & Troseth, 2014; Troseth, 2003b). Thus the deficit appears to be the result of several converging factors. A better understanding of the size of the deficit and the moderators associated with it can help to elucidate the contexts in which these factors tend to converge to result in the most significant deficits in learning.

Potential Moderators

Age. Many authors agree that the video deficit is present in children between the ages of 12 and 21 months. Most commonly, authors writing about the video deficit have chosen children’s third birthday as a reference point for when the video deficit is resolved or dramatically reduced for many tasks (Calvert & Richards, 2014; Kirkorian, Wartella, & Anderson, 2008; Roseberry et al., 2014; Sage & Baldwin, 2015), although earlier (Calvert & Wartella, 2014; Uhls, Michikyan, Morris, & Garcia, 2014) and later (Barr, 2010; Moser et al., 2015) estimates have also been used. These estimates have sometimes been based on developmental studies showing the video deficit for a particular task is reduced in older age groups (e.g., Dickerson, Gerhardstein, Zack, & Barr, 2012; Schmitt & Anderson, 2002). Some authors have suggested that the deficit peaks sometime during the first half of the second year (e.g., 15 months) and then may slowly diminish with age (Barr, 2010; Dickerson et al., 2013).

Most authors do not mention a lower age limit, although some suggest that children under 6 or 12 months may not display the deficit (Barr, 2010; Barr, Muentener, & Garcia, 2007; Krcmar, 2010).

The video deficit could be expected to lessen with age because of multiple developmental factors, which, combined with task demands, result in differential performance as children grow older. These factors may include development of working memory (Kirkorian, 2018), less reliance on social cues for learning, increased experience with video as relevant and meaningful to real life (Kirkorian & Choi, 2017; Troseth, 2003a, 2010; Troseth et al., 2007), and increased practice with symbolic thinking and linking video and reality (Strouse & Troseth, 2014). Children's development and experience may support them in learning from video, but real differences between video and face-to-face learning situations could also result in the persistence of the deficit. The video deficit may continue to occur when tasks are difficult but surmountable, because in these cases the extra working memory demands of learning from video or the mismatch in retrieval cues between video learning and in vivo applications may result in observable differences in learning and transfer (Barr, 2010; Kirkorian, 2018). Conceptual barriers may also still be present for older individuals. For example, some authors have argued that adults may learn less from digital than print texts because they perceive that screens are intended for shallow learning and may therefore invest fewer cognitive resources in the process (Ackerman & Goldsmith, 2011).

In the current study we test the continuous effect of age across the full range of our sample, because we expect that the developmental processes relevant to the deficit occur gradually over time. In addition, we investigate the moderating role of age by splitting our sample at 36 months, the age most commonly referenced as an upper limit for the video deficit. This enables us to estimate the size of the deficit both below and above this threshold.

Live Video. Video that is displayed on a screen "live," or as it is recorded, provides an opportunity for researchers to match video and face-to-face presentations. Face-to-face demonstrations involve small variations in delivery--perhaps researchers vary in speed, eye contact, facial expressions, or stumble on a word in the script. Distractions in the room may momentarily disrupt information delivery. The use of live video (e.g., closed-circuit video or video chat) allows for researchers to better match real world conditions, while still delivering information through the video medium. Because of the similarity in information delivery between live video feeds and face-to-face situations, it would be reasonable to expect learning in these conditions to be more similar than when pre-recorded videos are used.

In addition, when a person appears on bidirectional live video (e.g., video chat) they have the opportunity to respond contingently to the viewer, as they generally would if they were face to face. Contingency between a teacher and learner is one of several cues that may signal to children that information is being intentionally conveyed and is pedagogical in nature (Gergely et al., 2007; Sage & Baldwin, 2011). Closed-circuit video and video chat have been used as one way of studying the role of social cues as a mechanism for the video deficit (Myers et al., 2017; Myers et al., 2018; Nielsen et al., 2008; Roseberry et al, 2014; et al., 2018; Troseth et al., 2006; Troseth et al., 2018). Some closed-circuit and video chat studies have found that socially-contingent video supported children's learning (Myers et al., 2017; Nielsen et al., 2008; Roseberry et al, 2014; Troseth et al., 2006), and one set of pediatric recommendations listed video chat as an exception to their no-screens guideline (Council on Communications and Media, 2016). However, several intentional design features of these studies resulted in mismatches in the social support provided between conditions (for a more detailed explanation see Strouse et al., 2018; Strouse, 2019), and more recent studies have found mixed or null effects (Myers et al., 2017; Myers et al., 2018; Strouse et al., 2018; Troseth et al., 2018). Meta-analysis provides another lens to examine whether using live video to closely match information delivery across video and face-to-face conditions results in a reduction of the size of the video deficit.

Learning Domain. Several authors have argued that task complexity differences result in differences in the size of the deficit (Barr, 2010; Kirkorian, 2018; Krcmar, 2010). For example, tasks that are longer, include more distractors, or require memory updating may require more working memory, resulting in cognitive overload in video situations (Barr, 2010). Perceptual, social, and linguistic cues may act as cues for later retrieval, and when these cues are missing or mismatched between video learning and later in vivo tasks, a greater deficit may be apparent (Barr, 2010). Additionally, some tasks require nuanced conceptual understanding of video and others do not (Krcmar, 2010; Troseth, 2010). For example, if children simply fail to realize that an event is occurring on video, or that video is any different than reality, they may use the video information readily -- no dual representation required (Krcmar, 2010; Troseth & DeLoache, 1998).

Video deficit research has largely occurred within three domains of learning: language learning, imitation, and object retrieval. Language learning and imitation are tasks in which children often rely on social cue information (Baldwin & Moses, 2001; Barr, 2010). For example, infants learn more words when parents are socially responsive (Tamis-LeMonda et al., 2014), and toddlers use social cues such as a speaker’s gaze to help them map new labels to new objects (Baron-Cohen, Baldwin, & Crowson, 1997). Children are more likely to imitate actions they believe are intentional than actions that are accidental or they believe were not intended to be taught (e.g., Carpenter, Akhtar, & Tomasello, 1998), which may be difficult to determine when social cues are reduced. Matching retrieval cues are also important for successful imitations; for example, imitation is more likely when the learning and testing objects and environments are matched (Barnat, Klein, & Meltzoff, 1996). Therefore, the social and perceptual mechanisms for the video deficit are likely to apply to language and imitation tasks.

Object retrieval tasks, however, may incorporate different challenges, especially across repeated trials. In object retrieval tasks, children are typically shown the location of a hidden object on video, then asked to retrieve the object in a comparable real-life setting. For example, they may be shown a video of a person hiding an object in a room, then asked to find the object in the real room (Schmitt & Anderson, 2002; Troseth, 2003a, 2003b; Troseth & DeLoache, 1998). On subsequent trials, children are shown a video of the object being hidden in a new location in the same room, and then given a chance to search. Success on these trials depends on dual representation (understanding that the event on video is both occurring on video *and* representing a toy being hidden in a real, adjacent room) as well as a conceptual understanding that the video gives relevant and meaningful information about the current state of the world (Kirkorian et al., 2016; Troseth, 2010). Children must update the memory formed from direct experience (where they found the object last) with the memory from the most recent video, and choose to search based on the video information (Kirkorian et al., 2016; Troseth, 2010). This may be especially difficult because perceptual information available in children's memory of the real world (where they last found the object) may serve as a stronger retrieval clue match during their real-world search than information from their memory of the video (Kirkorian et al., 2016). Because of these additional challenges to using video information to successfully find hidden objects across multiple trials, object retrieval from video information may require a more nuanced conceptual understanding of the video medium, and the video deficit may be more apparent in this than in other domains.

The Current Study

In the current study, we review quasi-experimental and experimental studies of learning in video and face-to-face situations with participants between the ages of 0 and 6 years of age. We use meta-analytic techniques to examine the average size of the video deficit across the literature, and investigate the moderating role of age, the use of live video, and learning domain in the size of the video deficit being reported.

Methods

Inclusion and Exclusion Criteria

Studies. Eligible studies used experimental or quasi-experimental designs for comparison of children's learning of content presented via video versus the same content presented face-to-face (live). Reports needed to include data that permitted calculation of a numeric effect size for at least one eligible outcome variable (or, necessary data was retrievable through contact with the author). Eligible studies were conducted in any country, but were reported in English. No restrictions were placed on publication status or date.

Participants. Eligible studies included at least one calculable effect size for participants who were on average 6 years of age (83.9 months) and younger and not selected based on having a developmental delay. In this literature, cross-sectional developmental studies that include multiple age groups are common. In this case, eligibility was determined separately for each age group. All participants in eligible groups needed to be below 8 years of age, with an average age of 6 years or younger. Each group (video and live) needed to include at least 6 participants.

Interventions. Each study's video condition needed to include some presentation of information, displayed within the video itself, with the expectation that children would learn, apply, or otherwise use the presented content. No restrictions were placed on the type of content delivered, but similar instructional content in similar dosage needed to be delivered in both video and live conditions. Comparison groups that did not receive any instruction (e.g., no-exposure groups) or received different instructional content were not eligible.

Video was defined as a visual stimulus displayed on a screen, with or without an audio component, that played without ongoing input from the viewer. Computer and touchscreen games were excluded on the basis that without user input, they would not advance. Live videos (including closed-circuit videos and video chat) were included; these videos may adapt based on viewer input but would not stop or cease to function if the viewer did not respond.

Outcomes. Eligible studies included at least one outcome designed to measure children's learning, application, or use of the delivered instructional content. Learning was defined broadly as any skill, knowledge or new information (procedural or factual) acquired as a result of the intervention. Some writers have focused on transfer of learning across dimensions (e.g., from a 2D screen to a 3D object) when discussing the video deficit as well as learning from other media (e.g., books, touchscreens; the 'transfer deficit;' Barr, 2010; 2013). However, we opted to use the original definition of the video deficit (Anderson & Pempek, 2005), and did not require eligible studies' learning outcomes to require transfer. That is, the learning outcome could be tested on the same screen as it was taught.

Search and Study Selection

Our comprehensive, systematic search strategy included keyword searching, evaluation of references included in eligible articles and review articles, searching of conference proceedings, personal contact with authors working in the field, and hand searching of any journals that included four or more eligible studies. Search terms included the exact phrase, “video deficit,” or combinations of: (video, television, dvd, tv, program*, educational media, baby media) AND (children, preschoolers, toddlers, infants, babies) AND (learn*, understand*). To provide coverage of all four major disciplines relevant to this area of research and its relevant grey literature, we first searched PsycInfo, Communication Abstracts, Communication and Mass Media Complete, ERIC, Medline, and Google Scholar. We then searched the reference lists of review articles and eligible articles identified through these searches. Next we searched conference proceedings of the Society for Research in Child Development and the International Conference on Infant Studies (note that Communication and Mass Media Complete, ERIC, and Google Scholar also index conference proceedings, which were included in the initial search). Finally, we contacted authors and searched journals that occurred more than four times in our list of eligible reports.

Coding and Creation of Effect Sizes

Effect size selection. In this body of literature, it was common for one written report to include the results of several independent experimental or quasi-experimental studies (e.g., reported as Study 1, Study 2, etc.). It was also common for reports to include cross-sectional developmental designs in which independent samples of participants were recruited at different ages. Thus each written report had the potential to yield several independent effect sizes based on unique samples of participants. Because these effect sizes were calculated from non-overlapping samples, they were considered independent and appropriate for inclusion in a single analysis (Lipsey & Wilson, 2001).

In the case that a report included multiple independent live and video conditions, the first author followed the following procedure to create effect sizes for inclusion (questions were resolved via discussion with the second author): First, groups with the procedures that most closely matched each other were paired together. For example, Troseth and colleagues (2018) reported live and video conditions in which the presenter was responsive to the child as well as live and video conditions in which the presenter was unresponsive. In this case, two effect sizes were computed, one for the responsive conditions and one for the unresponsive conditions. Next, in the case that the report included an uneven number of video and live conditions, the conditions that were the best procedural match for each other were chosen. Finally, in the rare case that there was not a clear procedural match, conditions that were the focal concern of the study or had the most complete data were chosen.

Because multiple outcomes per study occurred rarely in this literature, when multiple outcomes were provided one was selected per study, to avoid dependency in the data. The outcome that most closely measured what was being taught in the treatment was used. If multiple outcomes were equal in match, the outcome that was the focal concern of the study, the most complete, or the most similar to other studies was chosen.

Coding. All study variables were coded by two independent raters. Variables coded for each written report (N = 59) included year of publication (ICC r = .997), the country in which the study was conducted (reduced to USA and Canada vs. other, kappa = 1), and the type of publication (peer-reviewed versus other, kappa = .813). Variables coded for each effect size (N = 122) included the research design (repeated-measures vs. independent groups, kappa = .709), the average age of children in months (ICC r = .999), whether live video was used (kappa = .961), and the learning domain of the outcome measure being reported (kappa = .885). Outcome measures were categorized as imitation (e.g., assembling toy, using a tool), language learning (e.g., choice of a labeled object), object retrieval (e.g., searching for a hidden object) or other (e.g., comprehension tests, recall questions). All coding discrepancies were resolved by the first author.

Effect size calculation. To compute effect size estimates and standard errors, data was entered into the Comprehensive Meta-Analysis (CMA) program (Borenstein, Hedges, Higgins, & Rothstein, 2005). CMA uses the effect size estimates and standard error calculations detailed in Hedges and Olkin (1985) that adjust for the small sample sizes common in this literature body. The majority of effect sizes (87%) were calculated based on means and standard deviations for each condition. The remaining effect sizes were calculated by entering the number of children succeeding or failing at a task (8%), a *t* value and sample size (4%), or a chi-square value and sample size (1%). All effect sizes were coded so that a positive effect indicated that participants performed better in the video condition and a negative number indicated an advantage to the live condition (i.e., a video deficit). In some cases, estimation of values was necessary, such as estimating sample means by measuring bar graphs or estimating group sizes by dividing the total sample by the number of groups. Exploratory analyses revealed no relation between effect size and estimation of group means or standard deviations. Studies with estimated group *n*s reported, on average, significantly stronger effects than the rest of the sample. However, this variable was confounded with age (estimated M = 23.32 months, not estimated M = 30.83 months, $t(53.234) = 2.66$, $p = .010$). Sensitivity analyses excluding studies where we had to estimate group sizes resulted in substantively identical results to those reported in this paper.

All data used for effect size calculations was double-coded after a time delay by the first author to confirm calculations, including any estimations, were consistent and accurate. A research assistant then coded 22% of the sample (n = 26), and showed acceptable reliability with the first author (ICC r = .99).

The CMA program was used to make appropriate adjustments for non-independence in within-subject designs. These adjustments require estimation of inter-sample correlations. Where not reported, inter-sample correlations were estimated as r = 0.5, based on correlations reported by Kendeou, Bohn-Gettler, White, & van den Broek (2008) for children's comprehension of material presented in different formats. Sensitivity analyses were conducted with effect sizes calculated using estimates of r = 0.8 and r = 0.2; results were substantively identical.

Analysis Plan

Once the data were reduced to a set of independent effect sizes and potential moderator variables, descriptive statistics were used to examine the distribution of effect sizes and of potential moderators. A series of random-effects meta-regression models were then run in the R package metafor (Viechtbauer, 2010). The first null model included no moderators and returned the overall, weighted mean effect size. Then, models examining study logistics (peer-reviewed, year published, repeated-measures), participant age, use of live video, and learning domain (imitation, language learning, object retrieval, other) were run to describe the relation between these potential moderators and the weighted average effect size. Finally, the possibility of publication bias was examined through standard methods (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Results

Literature Search

Our initial keyword searches resulted in 14,548 database hits (including duplicates), which the first author screened based on the eligibility criteria described above and erring on the side of inclusiveness. Because the initial search was so broad, the vast majority of reports were ineligible, most often because they used video as a stimulus but did not compare video versus live conditions, or because they included only children older than six years; these characteristics were clear from the abstracts. The initial search resulted in a list of 144 potentially eligible sources, which were then obtained in full text and screened in more detail. The first author then 1) searched the reference lists of 12 relevant reviews identified during the initial search and all articles deemed eligible, 2) paged through conference proceedings, 3) requested contributions from researchers who authored or co-authored 4 or more eligible reports, and 4) hand-searched the contents of four journals which had each contributed at least 4 reports. All together, these additional searches contributed another 66 potentially eligible reports, yielding a total of 210 potentially eligible reports. Final searches were completed on June 5, 2018.

We then attempted to attain full copies of all 210 potentially eligible reports. Authors were contacted as needed to request reports or missing information. When data was reported in multiple locations, the published version was preferred. If no version was published, the most complete or most recent version was used. Fifty-nine written reports (122 independent effect sizes) were identified for final inclusion. See Figure 1 for details of the search process and exclusions.

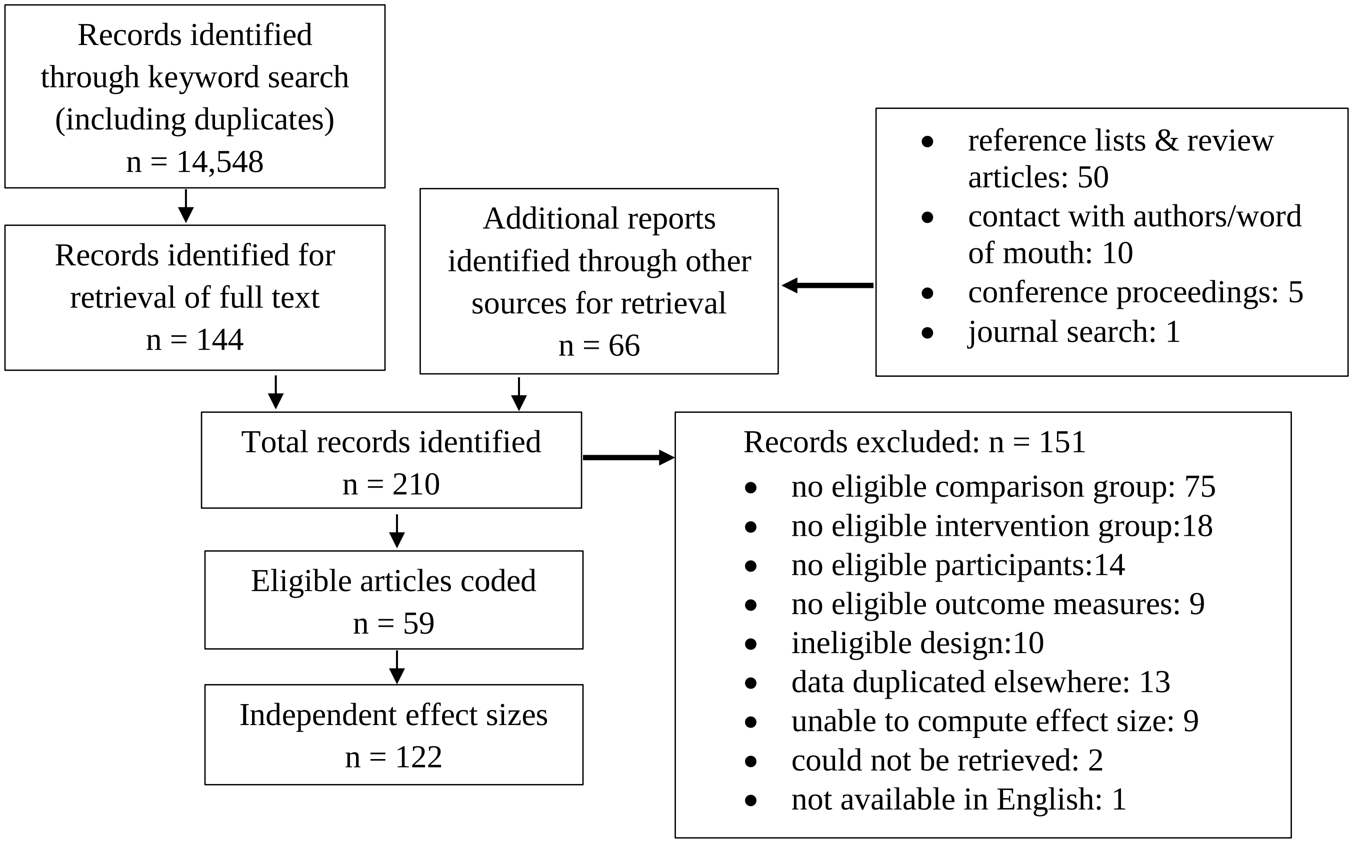

Figure 1. Literature search and screening

Study Characteristics

The 59 eligible reports (See Supplemental Material) were published between 1994 and 2018. The majority were published in peer-reviewed journals (73%) and used between-subjects designs (83%). Of the 122 included independent samples, 27% had an average age greater than 36 months, and the mean age of the samples was 29.48 months (*SD* = 16.92). Twelve percent of the effect sizes were from experiments using live video. The majority (60%) of effects sizes were based on imitation tasks, with the rest approximately equally distributed between language learning (17%), object retrieval (13%), and other (10%).

Overall Weighted Average Effect Size

The 122 included effect sizes ranged from -3.75 to 1.18 and displayed a somewhat negatively skewed distribution (see Figure 2). Sensitivity analyses with distributions Windsorized to 3 *SD* and 2 *SD* produced substantially identical results to those reported here. The weighted average effect size $\bar{g}$ was equal to -0.53, 95% CI [-0.66, -0.41], indicating an average video deficit of approximately one half of a standard deviation. The effect sizes demonstrated significant variability (*tau-square* = 0.33, $Q(121) = 404.95$, $p < .0001$), supporting the examination of potentially moderating variables.

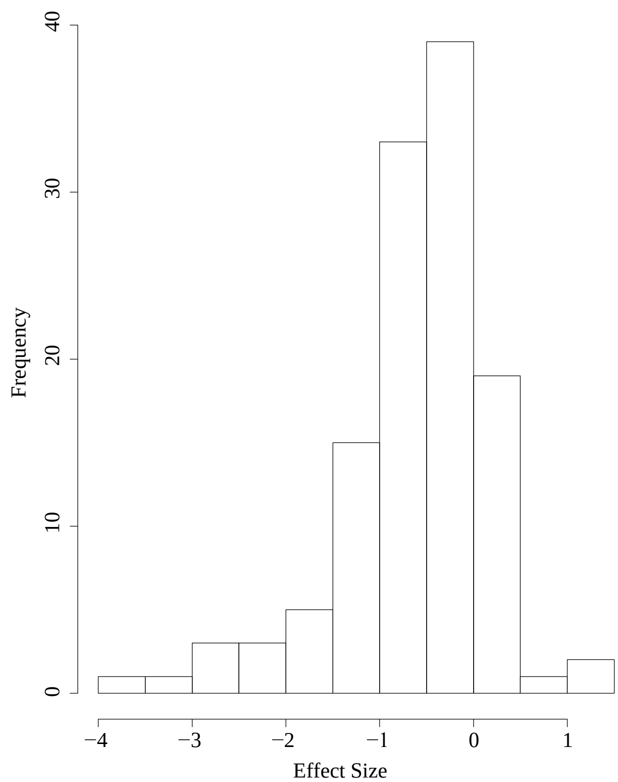

Figure 2. Distribution of effect sizes

Moderator Analyses

Random-effects meta-regression models were used to test potential moderating variables (see Table 1 for full results). Moderators included study logistics, age, the use of live video, and learning domain.

Table 1

Meta-regression Coefficients (bs)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	-.53**	-61.91**	-0.79**	-0.64**	-0.49**	-0.97**
Peer Review		-0.25				
Year		0.03**				
Repeated-Measures Design		0.45**				
Age (cont.)			0.01*			
Over 36mo				0.40**		
Live Video					-0.33	
Imitate						0.40*
Language Learning						0.68**
Other						0.77**
tau^2	0.33	0.27	0.31	0.29	0.32	0.29

Note. * $p < .05$, ** $p < .01$

Study logistics. The study logistics model included dichotomous codes for peer review and repeated-measures study design, and a continuous indicator of publication year. There was no difference in the effect sizes reported in peer-reviewed studies versus those that were not peer reviewed. Newer studies (*b* = 0.03, 95% CI 0.01, 0.05]) and those using repeated-measures designs (*b* = 0.45, 95% CI [0.13, 0.78]) reported a significantly smaller video deficit than their counterparts.

Age. We tested age as a moderator in two different ways. First, we included age in months as a continuous variable. The resulting model supported the presence of a small effect for age with studies using older participants reporting smaller video deficits (*b* = 0.01, 95% CI [0.002, 0.02]). That is, with each additional month of age, the deficit decreased by .01 standard deviations.

Because of the frequent reference in the literature to 36 months as an age at which the video deficit is greatly reduced or no longer observed, we also compared the size of the deficit for children above and below this threshold. Effect sizes were coded into younger (less than 36 months) and older (36 months or greater) samples based on the average reported age of participants. Studies with participants 36 months or older reported average effect sizes significantly closer to zero (less video deficit) than those with participants younger than 36 months (*b* = 0.40, 95% CI [0.13, 0.66]). Further analysis revealed that, regardless of age, the weighted average effect size was significantly different from zero (see Figure 3). Specifically, for studies including younger participants, the weighted mean effect size $\bar{g}$ was equal to -0.63, 95% CI[-0.77, -0.50]). For studies including older participants, the weighted mean effect size $\bar{g}$ was equal to -0.25, 95% CI[-0.50, -0.01]).

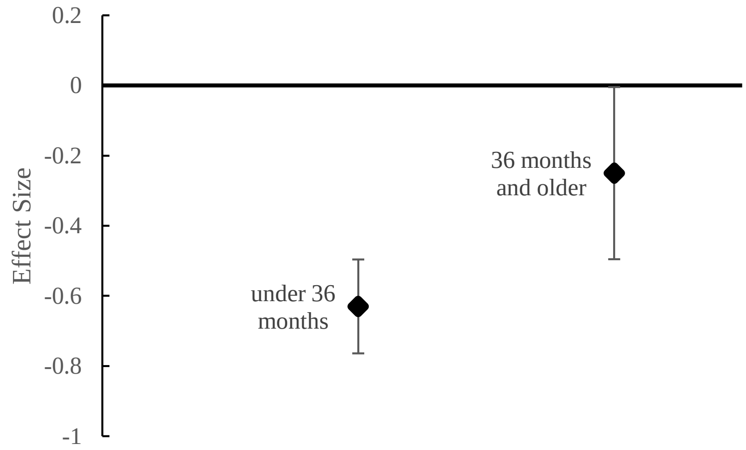

Figure 3. Weighted average effect size and 95% CI by age group

Live video. The next model tested whether there was a difference in the size of the video deficit for studies that used live videos in comparison to those in which the video was pre-recorded. There was no significant difference (*b* = -0.33, 95% CI [-0.70, 0.03], p < .10), and both types of studies reported weighted average effect sizes significantly less than zero (live video $\bar{g}$ = -0.82, 95% CI [-1.17, -0.48]; pre-recorded $\bar{g}$ = -0.49, 95% CI [-0.62, -0.36]). Because the direction of the effect was unexpected based on prior literature, we conducted exploratory follow-up analyses to examine whether the presence of a two-way feed was associated with a different size of effect than a one-way feed. Within the subgroup of studies using live video, we coded each effect size as either resulting from the use of a single-direction video feed (i.e., the child viewed a live video feed of the researcher, but there was no camera pointed at the child to provide information back to the researcher), or a bidirectional feed (i.e., both the child and researcher had information about each other; reliability, kappa = .84, one disagreement resolved through discussion). Studies using bidirectional feeds reported effect sizes closer to zero, representing smaller video deficits: *b* = .89, 95% CI[.37, 1.42].

Learning domain. A final model tested for differences in effect sizes related to the content domain (imitation, language learning, object retrieval, other). Object retrieval tasks, which were predicted to have the largest video deficit, were used as the reference category. Compared to object retrieval tasks, studies using imitation tasks (b = 0.40, 95% CI [0.04, 0.76]), language learning tasks (b = 0.68, 95% CI [0.26, 1.11]), and other tasks (b = 0.77, 95% CI [0.29, 1.25]) all reported effect sizes closer to zero (i.e. smaller video deficits). Additional models rotating the referent produced no additional significant differences between groups. Further analyses revealed that all groups except other skills reported weighted average effect sizes significantly less than zero (See Figure 4; imitation, $\bar{g}$ = -0.58, 95% CI -[-0.76, -0.41]; language learning, $\bar{g}$ = -0.27, 95% CI [-0.41, -0.14]; object retrieval, $\bar{g}$ = -1.00, 95% CI [-1.40, -0.60]; other, $\bar{g}$ = -0.16, 95% CI [-0.40, 0.08]).

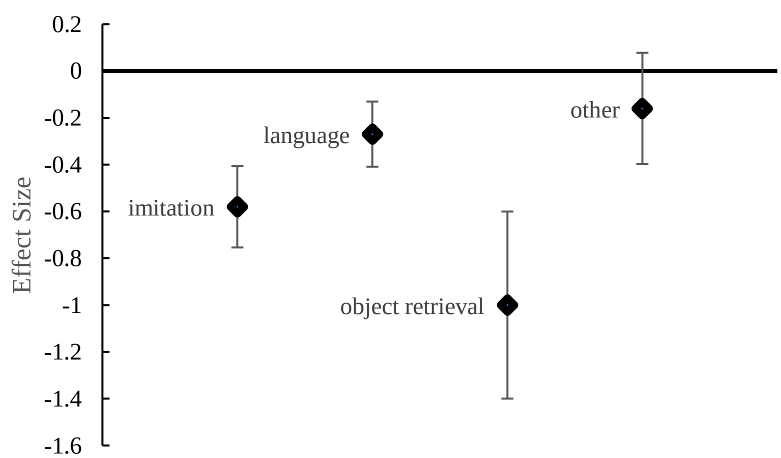

Figure 4. Weighted average effect size and 95% CI by learning domain

Publication Bias Analysis

A funnel plot and Egger's regression test revealed a significant relation between standard error and effect size, such that smaller (less precise) studies were more likely to report larger magnitude effect sizes (Egger's z = -7.84 p < .001, Figure 5). In other words, there appeared to be some "missing" small studies with smaller effect sizes. A trim and fill procedure suggested that, if there were missing studies, the inclusion of these studies would not substantially change the results. Specifically, the trim and fill procedure provided a weighted average effect size estimate of $\bar{g}$ = -0.47, 95% CI [-0.61, -0.33] after imputation of "missing" studies (Figure 5). In addition, a cumulative meta-analysis procedure indicated that the least precise studies did drag the overall weighted average effect size further from zero, but the average effect size was still significantly smaller than zero with the lower precision studies removed (e.g., with only the half of the sample with the smallest standard error included, $\bar{g}$ = -0.16, 95% CI [-0.27, -0.06], Figure 6).

Some researchers suggest trim and fill procedures do not always provide unbiased weighted average effect size estimates because they rely on the assumption that studies with small effects, rather than non-significant results, will be unpublished (Simonsohn, Nelson, & Simmons, 2014). To address this concern, we computed the inflation rate and r-index described by Schimmack (2016). The inflation rate represents the discrepancy between median observed power and the percentage of significant results. Our calculated inflation rate of 0.88% (median observed power = 35.19%, success rate = 36.07%) indicates little evidence of bias in the studies included in this analysis (Shakil & Schimmack, 2015). However, the r-index of 34.31% (median observed power – inflation rate) suggests that researchers in this field should consider power when designing future studies.

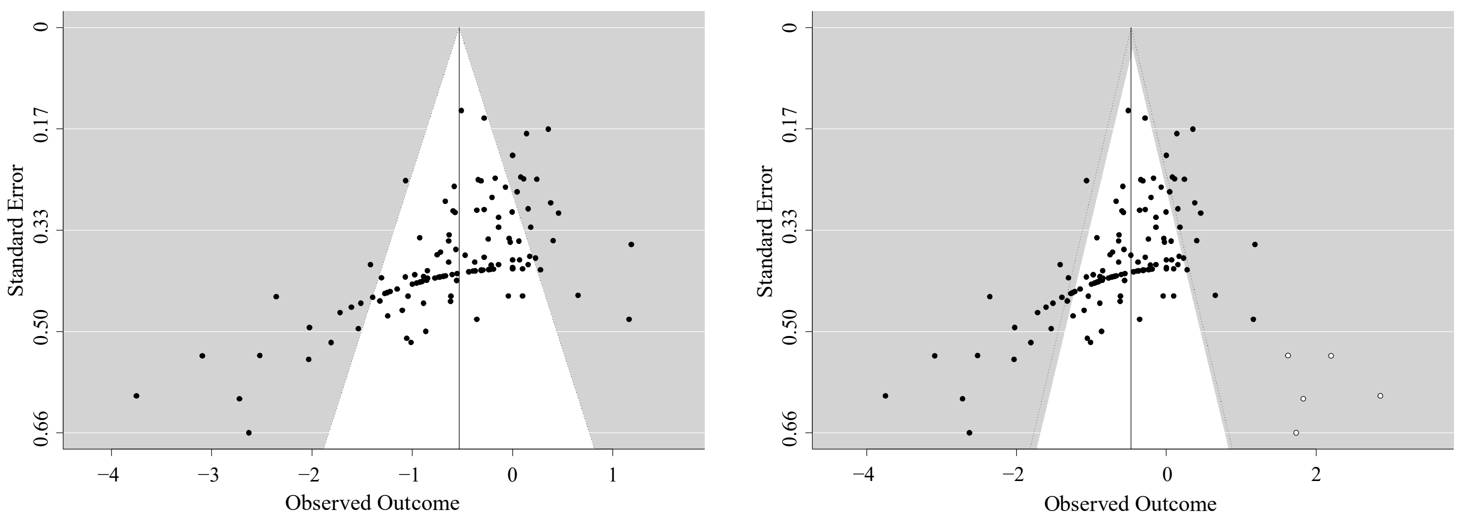

Figure 5. Original funnel plot(left) and trimmed and filled funnel plot (right)

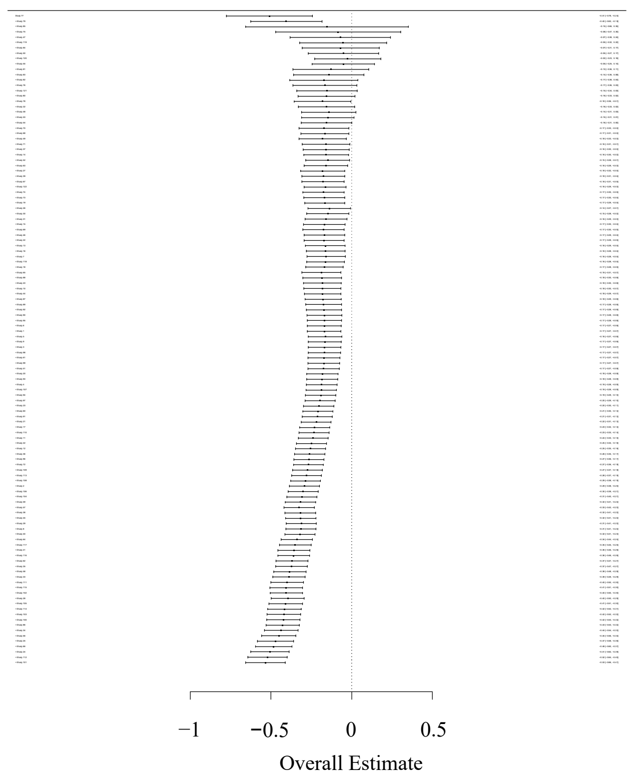

Figure 6. Cumulative forest plot, showing overall weighted average effect size as each individual study is added to the analysis in order from smallest SE to largest

It is also possible that the relation between effect size and standard error is an artifact of some confounding variable about the smallest studies (Borenstein et al., 2009). To explore this more fully, we checked for a relation between standard error and our significant moderators. There was no relation between standard error and age (r = -0.05, p = .575), but standard error did differ by learning domain. That is, studies in the two domains with the largest effect sizes also reported the largest standard errors, imitation, SE = .42, object retrieval, SE = .39, other, SE = .32, language, SE = .31. An ANOVA, $F(3,118) = 12.43$, $p < .001$, with Tukey post-hoc tests indicated that language tasks had smaller standard errors than imitation ($p < .001$) and object retrieval tasks (p = .028), and 'other' had smaller standard errors than imitation tasks ($p < .001$). Additional follow-up testing suggested, however, that even within domain, there was still a relation between effect size and study precision (Egger's regression tests: imitation z = -7.51, p < .001, language z = -0.43, p = .669, object retrieval z = -4.47, p < .001, other z = -2.93, p = .003). Therefore, it is unclear the extent to which publication bias, learning domain, or some other unmeasured confound is responsible for the association between effect size and standard error.

Discussion

Our meta-analysis results suggest that, on average and across ages, video types, and learning domains, children scored approximately half a standard deviation ($\bar{g}$ = -0.53) higher on tests of learning when they were taught the information face-to-face rather than watching the same information given on video. We also observed multiple ways in which these effects varied, especially by age and domain. Because moderator effects are correlational in nature (Lipsey, 2003), experimental research is needed to determine whether the effects we describe are causal. Publication bias and unmeasured confounds such as task difficulty, contextual supports for learning, or the way that learning was measured, could influence the average effects observed.

Moderators

Age. We found that the video deficit, or the difference in children's learning from live versus video presentations, slowly diminished with age, consistent with some predictions made in the literature (e.g., Dickerson et al, 2013). Slow reduction in the deficit with age is compatible with several theoretical mechanisms for the deficit, because development of working memory (Kirkorian, 2018), ability to transfer across contexts (Barr, 2013), and ability to use a wider variety of referential cues for learning (Akhtar & Tomasello, 2000) should occur with age. Children should also acquire more experience learning from video (Troseth et al., 2019). However, when considering children's performance across all tasks, we found a deficit of approximately one quarter of a standard deviation for children older than 3 years. This smaller magnitude deficit may result from a reduction in the number of converging factors contributing to the deficit at older ages, and possibly reflects a pattern in which the slow reduction in the deficit leads to it becoming insubstantial toward the upper limits of our age range. We did not have adequate data to determine the specific age at which the effect may disappear. Age is also confounded with task in this literature, so further research is needed to explore how age interacts with task demands.

Live video. Contrary to our expectation and some previous literature in which live video has been used to support children's learning (Myers et al., 2017; Nielsen et al., 2008; Roseberry et al, 2014; Troseth et al., 2006), the use of live video did not significantly moderate the average effect size in our analyses. Studies using both live ($\bar{g}$ = -0.82) and pre-recorded ($\bar{g}$ = -0.49) video reported deficits significantly different from zero. We explored the presence or absence of a bidirectional feed, through which the researcher could potentially react contingently to the viewing child's action, as a potential moderator. Studies with bidirectional feeds showed less deficit (effect sizes closer to zero) than studies with one-way feeds, indicating that children may be sensitive to two-way contingency within live video. However, this analysis should be interpreted with caution due to the large variability in effect sizes in this small subsample (n = 15), and confounding of learning domain and publication year with the bidirectional feed variable (see Supplemental Materials for details on how these were coded). In addition, our effect size coding procedure involved selecting study conditions for inclusion that provided the best procedural match with one another. This resulted in the formation of effect sizes using the pre-recorded condition rather than the live video condition in one study (Nielsen et al., 2008) and a live video condition that involved less rather than more training regarding its live status in another study (Troseth et al., 2006). Our selection criteria therefore resulted in a moderator analysis in which the live video conditions that involved less social support were included in these cases. The resulting lack of difference between live and pre-recorded video aligns with several recent live video studies, which have concluded that supportive scaffolding may help children to fully benefit from contingency offered through live video (Myers et al., 2018; Strouse et al., 2018).

Learning domain. The magnitude of the video deficit was largest, one full standard deviation in size, for object retrieval tasks. Object retrieval tasks, as they appear in the current literature, may be a scenario where the multiple causes of the video deficit converge to result in an especially large deficit. These tasks may be particularly difficult for children when information is presented on video, because after the first trial they also have direct experience with the hiding locations that competes with information presented on video (Kirkorian et al., 2016). In these cases, children must realize that the video is intended to give them updated information about where the object is hidden, and must choose to use this new information rather than rely on where they last found the object (Kirkorian et al., 2016). Children's relatively low performance on this type of video task likely does not represent difficulty with spatial tasks generally, but rather with the availability of directly competing firsthand experience and video information. To make the choice to use the video information, children must conceptually understand the video as relevant, meaningful, and intended to represent the real world (i.e., dual representation, Troseth, 2010; Troseth & DeLoache, 1998). Therefore, video object retrieval tasks carry not only the cognitive demands of other video tasks, but also require a more mature understanding of video as a symbolic medium.

The magnitude of the deficit was smaller for language learning ($\bar{g}$ = -0.27) and imitation ($\bar{g}$ = -0.58) tasks. Children rely on social cues to support learning in these domains (Baldwin & Moses, 2001; Barr, 2010), and rely on similarities between learning and testing environments to support transfer and application of learned information (Barr, 2010). However, children may not need dual representation to solve simple language learning and imitations tasks (Krcmar, 2010; Troseth et al., 2019). For example, if infants do not notice that an action is occurring on screen, and think of it as if it were happening in vivo, they do not need to represent both the screen-based representation and the referent because they do not need to represent the screen-based representation at all. This may explain why 6-month-olds did not display the deficit when asked to imitate simple actions done with a puppet (Barr et al., 2007). The larger magnitude of the deficit for object retrieval tasks, paired with a smaller but still significant effect for imitation and language learning tasks, supports the idea that several mechanisms are at play in causing the deficit, such that even when dual representation is not required, social, perceptual, and contextual differences may still impact children's learning.

There was no evidence of a significant video deficit for studies in our *other* skills category, which included theory of mind, problem solving, and a variety of comprehension and recall tasks. This category was comprised of the smallest number of studies (n = 12), participants who were older than average, and tasks that were more variable and potentially less complex than the other categories. Success at the tasks in this category may not have relied on as many of the mechanisms relevant to the video deficit, as tasks were generally not set up to require transfer or dual representation, and children may not have been as dependent on social cues for success. However, tasks in this category may still be susceptible to the deficit. For example, video presentation of information could result in cognitive overload, or children may fail to devote cognitive resources to processing video information because they do not perceive its relevance. Further research is needed to determine whether the deficit is observed for any tasks in this group, and disambiguate reasons why the deficit may or may not be apparent.

Study logistics. Finally, our analyses revealed a *decline effect* (Pietschnig, Siegel, Nur Eder, & Gittler, 2017; Protzko & Schooler, 2017) in that less video deficit was reported in more recent studies. Decline effects may reflect an artificial inflation of the initial effect size due, for example, to the use of small samples and pressure to publish significant and surprising effects. In these cases, researchers attempting to replicate the effect over time find and publish smaller effects that better represent the true effect. It is also possible for decline effects to be observed when the necessary conditions to reproduce the effect are not well specified (Protzko & Schooler, 2017). Our results might represent an effect of one of these types.

An additional explanation is that a historical trend may contribute to the declining size of the reported video deficit over time (a “genuinely decreasing decline effect,” Protzko & Schooler, 2017). Children today are exposed to a wider variety of screen-based media than they were when this type of research began, and touchscreen devices allow for young children to interact contingently with devices in a much more accessible manner than the computers and mice of the 1990s and early 2000s (Hitlin, 2018). Early studies reported that children exposed to videos of themselves at home or on store security cameras were more likely to apply information from video to in-person tasks than children who did not see themselves on video (Troseth, 2003b; Troseth et al., 2007). Today, nearly all children in the U.S. have devices in the home that allow them to record and replay videos of themselves and family members (Hitlin, 2018; Rideout, 2017) As a result of these experiences, children may develop a more robust concept of how video might be used to provide information that is meaningful in the real world at younger ages than in prior decades.

Therefore, it is also important to explore whether the same deficit and moderators we have identified here are also associated with children's learning from other two-dimensional media such as touchscreen games and print books, or whether the type of two-dimensional media used is an additional moderator. As outlined by Barr (2010), some factors relevant to the video deficit likely continue to operate in other media contexts. For example, perceptual and contextual mismatches may apply whenever the learning and application contexts differ. However, other factors may not: touchscreen tablets can provide contingent feedback to children that can direct their attention and support their learning (Kirkorian, 2018; Troseth et al., 2019), and co-readers can provide social supports for learning from print books. There may be a general deficit in transfer that applies across media, but unique attributes of each medium may serve to moderate the size of the effect.

Limitations and Future Directions

A meta-analysis is only as good as the body of literature which it summarizes (Lipsey & Wilson, 2001), and the current synthesis is limited by several features of the existing research on the video deficit. First, several variables, if adequate variation had occurred in the literature, might have improved the current moderator analyses. The most pressing of these was that we did not have a sufficient number of effect sizes available to test for lower and upper age limits of the video deficit. Some have suggested that children under 6 or 12 months may not display the deficit (Barr, 2010; Barr et al., 2007; Krcmar, 2010). We coded just seven eligible effect sizes for children below the age of 12 months, and zero eligible effect sizes for children below the age of 6 months. Of the seven effect sizes observed, four reported higher performance in the face-to-face condition, two reported higher performance in the video condition, and one reported equivalent performance. With the limited amount of information available, we were unable to estimate whether the size of the effect in this age range was significantly different from, or whether it may be diminished in comparison to, the effect observed for children over 12 months of age. Similarly, we had a small proportion of our total number of effect sizes coded for children over 36 months (33 out of 122), which left us unable to address age effects or task-related confounds within the older group.

Another potentially interesting moderator involves the production values of the videos used. The goal of the current study was to address differences in children's learning from presentations that only differed in the format through which information was delivered, and to provide this level of match the studies reported here used videos that were almost entirely researcher-created (through direct filming of the same event that occurred live). Only one qualifying study used commercially-produced video. Some have argued that production values are important for capturing children's interest and engagement, and that poor production values could be partially responsible for instances in which children display low levels of learning from video (Barr & Hayne, 1999; Krcmar & Cingel, 2019). Similarly, the lab-based environments typical in many video deficit studies may be less relevant to children's real-world learning than real-world contexts (Krcmar, 2010). Many studies using commercially-produced videos, often in naturalistic environments, have shown that children display significant learning gains (e.g., Linebarger, Kosanic, Greenwood, & Doku, 2004; Rice, Huston, Truglio, & Wright, 1990; Zill, 2001). However, these studies have often involved comparison of children's learning from commercially-produced videos with no exposure control groups or groups that watched a program with different content, and were therefore ineligible for the current study. These studies are able to show that young children can and do learn from commercially-produced videos, but are unable to provide information about how learning from these videos compares with learning from face-to-face presentations. Future research should address the applicability of the video deficit reported here to more commercial uses of video.

Finally, we are limited by the quality of the studies in the existing literature. Our eligibility criteria resulted in the inclusion of only quasi-experimental and experimental studies with similar content delivered in live and video formats. As such, some level of match between the conditions was required for inclusion. However, between-subjects designs reported larger deficits than repeated-measures designs, which may indicate that random assignment did not entirely eliminate individual differences between conditions in these studies (as can happen with small samples). This research body is comprised primarily of small studies with low power and unstandardized outcome measures that have not been pre-registered. Inconsistency in reporting meant we had to estimate information such as the number of participants per group where this information was not included and authors did not or could not provide it. Although our inflation factor suggests little bias due to lack of publication of non-significant results, our other publication bias analyses suggest there may be some studies that reported small effects that never made it out of the researchers' file drawers, or that some confound with small sample sizes exists. Therefore, it is likely that some bias is also present in this field and affects the estimate of average effect size provided by meta-analytic synthesis. We hope future researchers will acknowledge and work to correct these limitations to the research body, especially using methods such as preregistration to reduce potential bias.

As an additional note, two of our weighted average effect sizes and one estimate of the overall weighted average effect size adjusted for potential publication bias, while statistically significant, were small in magnitude. While these effect sizes exceed the minimum guidelines for interpretation put forward by other statisticians (Cohen, 1988, Lipsey, 1998), Ferguson (2009) argued that effect sizes smaller than .41 should not be interpreted as practically significant unless outcome measures were highly valid or studies were rigorously controlled. In particular, Ferguson (2009) suggests that small effect sizes may be more driven by "noise" in the data than by true effects. Although this literature is largely comprised of controlled, lab-based experimental designs, studies have also tended to use small samples with low power. There is evidence that suggests there may be publication bias or small sample confounds that make it especially appropriate to proceed with caution when interpreting effects in this range.

Research with children's learning is inherently messy, with multiple variables interacting, which can result in smaller effect sizes for the isolation of one variable. Even the smallest of our significant effect sizes, which represent a comparison of children's learning of similar content delivered in two different ways, are in line with what is typical when two similar educational interventions are compared (mean effect size .2-.3 across 76 meta-analyses; Hill, Bloom, Black,

& Lipsey, 2008) and they are also typical of effect sizes reported and interpreted by recent authors of meta-analyses in similar content areas or with children of similar age (e.g., Dowdall et al., 2019; Groh & Narayan, 2019; Ulferts, Wolf, & Anders, 2019). If these reports are indications that true effects in the .2-.4 range are desirable topics of investigation and interpretation, future researchers will need to address sources of "noise" that limit the current interpretability of effects in this range. At this point, researchers, practitioners, and policy-makers will need to decide for themselves what constitutes an interpretable and practically significant effect in their context.

Implications

Regardless of how the smaller weighted average effect sizes reported here are interpreted, these analyses contribute to understanding the multiple converging factors associated with the video deficit. Our analyses demonstrated that a decrease in the magnitude of the video deficit occurred with age. With age and experience, children may develop better conceptual understanding of the role of video and its relevance in their lives. They may rely less on surface features as retrieval cues for transfer, and rely less on social contingencies for identifying pedagogical contexts. However, the video deficit may continue to appear across the lifespan when tasks are difficult, because older children and adults continue to be influenced by some the same factors implicated by video deficit researchers. For example, adults are better able to retrieve information from memory when learning and testing contexts are similar (Smith & Vela, 2001), and they use communicative pedagogical cues to draw inferences about what information should be learned (Shafto, Goodman, & Frank, 2012). Adults and older children may believe information presented on screens is "easier" and requires less mental effort than information presented in other contexts, resulting in lower effort and overconfident predictions about knowledge acquisition from videos and digital books (Ackerman & Goldsmith, 2011; Salomon, 1984). Therefore, even though the video deficit may be reduced at older ages, researchers interpreting performance on screen-based research tasks and educators developing curricular materials may wish to consider that adults and older children learning from video may still be subject to some of the same constraints as younger children.

Learning domain differences are also consistent with the video deficit resulting from multiple mechanisms. Domain differences may be the consequence of differences in task demands associated with the procedures typically used in these domains. For example, object retrieval tasks, in which children displayed the largest difference between learning from video and live presentations, were set up to require substantial conceptual understanding of how video could be used to provide information about the task across multiple trials. Imitation tasks, which often relied on children's memory of video information over a delay, may be more heavily influenced by the perceptual match between learning and test environments. Imitation and language learning tasks may more heavily rely on social cue information and the presence or absence of social contingency. Tasks also differed in the extent to which they required transfer of the information learned to new contexts. These task demands, and their association with the perceptual, social, and conceptual mechanisms proposed for the video deficit, are likely more important predictors of children's performance than the particular domain of study itself. Future researchers may consider whether changing the demands typically associated with a given domain, such as adding competing information across multiple trials to imitation or word learning studies, would impact children's performance in these domains.

In addition, the consistency of the deficit across both live video and pre-recorded contexts, along with the confounds present in our exploratory analysis of bidirectional feeds, suggests that video deficit research may not currently provide support for differential policy guidelines based on this feature. Other research has highlighted potential benefits of involving young children in video chat: young children appear to feel secure in a new environment when a parent interacts with them through video chat (Tarasuik, Galligan, & Kaufman, 2011) and prefer and recognize adults they have interacted with through video chat over strangers (Myers et al., 2017). Thus, video chat may support social connections between children and people in remote locations. Toddlers also are reported to be quite engaged by people who respond contingently on video, which, if it serves to maintain children's interest and attention to information presented on screen, may lead to learning over time (Strouse et al., 2018). In addition, studies that included training experiences with video chat appear to show more benefits to learning (Nielsen et al., 2008; Troseth et al., 2006). More research is needed to determine the best social and contextual supports to optimize children's learning.

Relatedly, it should also be acknowledged that learning is not the only reason children interact with video. Parents report that they provide screen media access to their children because children enjoy it, and to help provide parents with time to accomplish other tasks (Cingel & Krcmar, 2013; Strouse, Newland, & Mourlam, 2019). However, inefficient learning is just one of multiple parent concerns about young children's exposure to screen media. For example, children may be exposed to advertisements, sexual and violent content, and gender and racial stereotypes through screen media (Rideout & Hamel, 2006; Rideout, 2017); and parents worry that screen media exposure may negatively affect their child's behavior, mood, social, and physical development (Bentley, Turner, & Jago, 2016; Decker et al., 2012). Pediatric guidelines cite concerns regarding the quality of existing children's media, displacement of social play and parent-child interaction, increased obesity resulting from exposure to advertising and eating while watching television, and disruption of sleep (Council on Communications and Media, 2016). Policymakers and caregivers should keep both educational opportunities and concerns in mind.

Summary

Across tasks and contexts, children ages 6 years and younger display approximately half of a standard deviation lower scores on learning outcomes after watching a video presentation than a face-to-face presentation of the same information. The size of the deficit decreases with age and is larger for children under 3 years than for children over 3 years. The deficit is largest for object retrieval tasks, when compared with imitation and language learning, although there is evidence of a small deficit in these domains as well. We had insufficient evidence to conclude there was a deficit for other tasks, such as comprehension and recall; more literature is needed in these areas. The pattern of results reported here supports the theoretical perspective that multiple mechanisms converge to cause the video deficit effect, including both developmental factors and the particular demands of the learning task.

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: on screen versus on paper. *Journal of Experimental Psychology: Applied, 17*, 18-32. doi: 10.1037/a0022086.

Akhtar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello, & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 115–135). New York, NY: Oxford University Press.

Anderson, D. R., & Pempek, T. A. (2005). Television and very young children. *American Behavioral Scientist, 48*, 505-522. doi: 10.1177/0002764204271506

Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 309–329. doi: 10.1111/1467-9507.00168

Barnat, S. B., Klein, P. J., & Meltzoff, A. N. (1996). Deferred imitation across changes in context and object: Memory and generalization in 14-month-old infants. *Infant Behavior and Development, 19,* 241-251. doi: 10.1016/S0163-6383(96)90023-5

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development, 68*, 48-57. doi: 10.1111/j.1467-8624.1997.tb01924.x

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review, 30,* 128-154. doi: 10.1016/j.dr.2010.03.001

Barr, R. (2013). Memory constraints on infant learning from picture books, television, and touchscreens. *Child Development Perspectives*, *7,* 205-210. doi: 10.1111/cdep.12041

Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development, 70,* 1067-1081. doi: 10.1111/1467-8624.00079

Barr, R., Moser, A., Rusnak, S., Zimmermann, L., Dickerson, K., Lee, H., & Gerhardstein, P. (2016). The impact of memory load and perceptual cues on puzzle learning by 24-month olds. *Developmental Psychobiology, 58,* 817–828. doi: 10.1002/dev.21450

Barr, R., Muentener, P., & Garcia, A. (2007). Age-related changes in deferred imitation from television by 6- to 18-month-olds. *Developmental Science, 10,* 910–921. doi: 10.1111/j.1467-7687.2007.00641.x

Bentley, G. F., Turner, K. M., & Jago, R. (2016). Mothers' views of their preschool child's screen-viewing behaviour: A qualitative study. *BMC Public Health, 16,* 718. doi: 10.1186/s12889-016-3440-z

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis 2.0. *Englewood, NJ: Biostat*.

Borenstein, M., Hedges, L. V., & Higgens, J. P. T. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley & Sons, Ltd.

Calvert, S. L., & Richards, M. N. (2014). Children's parasocial relationships. In A. B. Jordan & D. Romer (Eds.), Media and the well-being of children and adolescents (pp. 187-200). Oxford, UK: Oxford University Press.

Calvert, S. L., & Wartella, E. A. (2014). Children and electronic media. In E. T. Gershoff, R. S. Mistry, & D. A. Crosby (Eds). *Societal contexts of child development: Pathways of influence and implications for practice and policy* (pp. 175-187). New York, NY: Oxford University Press.

Canadian Paediatric Society (2017). Screen time and young children: Promoting health and development in a digital world. *Paediatrics & Child Health, 22*, 461–468. doi: 10.1093/pch/pxx123

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*, 315-330. doi: 10.1016/S0163-6383(98)90009-1

Carver, L. J., Meltzoff, A. N., & Dawson, G. (2006). Event-related potential (ERP) indices of infants' recognition of familiar and unfamiliar objects in two and three dimensions. *Developmental Science, 9*, 51-62. doi: 10.1111/j.1467-7687.2005.00463.x

Choi, K., Kirkorian, H., & Pempek, T. (2018). Understanding the transfer deficit: Contextual mismatch, proactive interference, and working memory affect toddlers' video-based transfer. *Child Development, 89,* 1378-1393. doi: 10.1111/cdev.12810

Cingel, D. P., & Krcmar, M. (2013). Predicting media use in very young children: The role of demographics and parent attitudes. *Communication Studies, 64*, 374-394. doi: 10.1080/10510974.2013.770408

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.) New York: Lawrence Erlbaum Associates. doi: 10.4324/9780203771587

Council on Communications and Media. (2016). Media and young minds. *Pediatrics*, 138. doi: 10.1542/peds.2016-2591

De Decker, E., De Craemer, M., De Bourdeaudhuij, I., Wijndaele, K., Duvinage, K., Koletzko, B Grammatikaki, E., Iotova, V., Usheva, N., Fernández-Alvira, J. M., Zych, K., Manios, Y., & Cardon, G. (2012). Influencing factors of screen time in preschool children: An exploration of parents' perceptions through focus groups in six European countries. *Obesity Reviews*, *13*, 75-84. doi: 10.1111/j.1467-789X.2011.00961.x

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood. A new imitation task. *Developmental Psychobiology, 55*, 719-732. doi: 10.1002/dev.21068

Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2019). Shared picture book reading interventions for child language development: A systematic review and meta-analysis. *Child Development.* doi: 10.1111/cdev.13225

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40,* 532-538. doi: 10.1037/a0015808

Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, *10*, 139-146. doi: 10.1111/j.1467-7687.2007.00576.x

Groh, A. M., & Narayan, A. J. (2019). Infant attachment insecurity and baseline physiological activity and physiological reactivity to interpersonal stress: A meta-analytic review. *Child Development, 90,* 679-693. doi: 10.1111/cdev.13205

Hedges, L. V., & Olkin, L. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2,* 172-177. doi: 10.1111/j.1750-8606.2008.00061.x

Hitlin, P. (2018, September 28). Internet, social media use, and device ownership in U.S. have plateaued after years of growth. Retrieved from http://www.pewresearch.org/facttank/2018/09/28/internet-social-media-use-and-device-ownership-in-u-s-have-plateaued-after-years-of-growth/

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*, 259-272. doi: 10.1111/j.1467-9817.2008.00370.x

Kirkorian, H. L. (2018). When and how do interactive digital media help children connect what they see on and off the screen? *Child Development Perspectives*, *12*, 210-214. doi: 10.1111/cdep.12290

Kirkorian, H. L., & Choi, K. (2017). Associations between toddlers' naturalistic media experience and observed learning from screens. *Infancy*, *22*, 271-277. doi: 10.1111/infa.12171

Kirkorian, H. L., Lavigne, H. J., Hanson, K. G., Troseth, G. L., Demers, L. B., & Anderson, D. R. (2016). Video deficit in toddlers' object retrieval: What eye movements reveal about online cognition. *Infancy*, *21*, 37–64. doi: 10.1111/infa.12102

Kirkorian, H. L., Wartella, E. A., & Anderson, D. R. (2008). Media and young children's learning. *The Future of Children, 18*, 39–61. doi: 10.1353/foc.0.0002

Krcmar, M. (2010). Assessing the research on media, cognitive development, and infants. *Journal of Children and Media*, *4*, 119-134. doi: 10.1080/17482791003629586

Krcmar, M., & Cingel, D. (2019). Do young children really learn best from the use of direct address in children's television? *Media Psychology*, *22*, 152-171. doi: 10.1080/15213269.2017.1361841

Kuhl, P. K. (2007). Is speech learning gated by the social brain? *Developmental Science, 10,* 110-120. doi: 10.1111/j.1467-7687.2007.00572.x

Linebarger, D. L., Kosanic, A. Z., Greenwood, C. R., & Doku, N. S. (2004). Effects of viewing the television program *Between the Lions* on the emergent literacy skills of young children. *Journal of Educational Psychology, 96,* 297-308. doi: 10.1037/0022-0663.96.2.297

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 39 – 68). Thousand Oaks, CA: Sage.

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science, 587,* 69–81. doi: 10.1177/0002716202250791

Lipsey, M. W., & Wilson, D. B. (2001). *Applied social research methods series: Vol. 49. Practical meta-analysis.* Thousand Oaks, CA: Sage Publications, Inc.

Moser, A., Zimmermann, L., Dickerson, K., Grenell, A., Barr, R., & Gerhardstein, P. (2015). They can interact, but can they learn? Toddlers' transfer learning from touchscreens and television. *Journal of Experimental Child Psychology*, *137,* 137–155. doi: 10.1016/j.jecp.2015.04.002

Myers, L. J., LeWitt, R. B., Gallo, R. E., & Maselli, N. M. (2017). Baby FaceTime: Can toddlers learn from online video chat? *Developmental Science*, *20*, e12430. doi: 10.1111/desc.12430

Myers, L. J., Crawford, E., Murphy, C., Aka-Ezoua, E., & Felix, C. (2018). Eyes in the room trump eyes on the screen: Effects of a responsive co-viewer on toddlers' responses to and learning from video chat. *Journal of Children and Media*, *12*, 275-294. doi: 10.1080/17482798.2018.1425889

Nielsen, M., Simcock, G., & Jenkins, L. (2008). The effect of social engagement on 24-month-olds' imitation from live and televised models. *Developmental Science*, *11*, 722–731. doi: 10.1111/j.1467-7687.2008.00722.x

Pietschnig, J., Siegel, M., Nur Eder, J. S., & Gittler, G. (2019). Effect declines are systematic, strong, and ubiquitous: A meta-meta-analysis of the decline effect in intelligence research. *Frontiers in Psychology, 19.* https://doi.org/10.3389/fpsyg.2019.02874

Protzko, J., & Schooler, J. W. (2017). Decline effects: Types, mechanisms, and personal reflections. In S. O. Lilienfeld & I. D. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 85-107). West Sussex, UK: Wiley Blackwell.

Rice, M. L., Huston, A. C., Truglio, R., & Wright, J. C. (1990). Words from "Sesame Street": Learning vocabulary while viewing. *Developmental Psychology, 26,* 421-428. doi: 10.1037/0012-1649.26.3.421

Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents.* Menlo Park, CA: The Henry J. Kaiser Family Foundation. Retrieved from https://www.kff.org/wp-content/uploads/2013/01/7500.pdf

Rideout, V. (2017). *The Common Sense census: Media use by kids age zero to eight.* San Francisco, CA: Common Sense Media. Retrieved from https://www.commonsensemedia.org/sites/default/files/uploads/research/csm_zerotoeight_fullreport_release_2.pdf

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85,* 956–970. doi: 10.1111/cdev.12166

Sage, K. D., & Baldwin, D. (2015). Children’s use of self-paced slideshows: An extension of the video deficit effect? *Journal of Research in Childhood Education*, *29*, 90–114. doi: 10.1080/02568543.2014.978919

Salomon, G. (1984). Television is “easy” and print is “tough:” The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*, 647-658. doi: 10.1037/0022-0663.76.4.647

Schmitt, K. L., & Anderson, D. R. (2002). Television and reality: Toddlers' use of visual information from video to guide behavior. *Media Psychology*, *4*, 51-76. doi: 10.1207/S1532785XMEP0401_03

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science, 7,* 341-351. doi: 10.1177/1745691612448481

Shakil, A., & Schimmack, U. (2015, August 5). Using the R-index to detect questionable research practices in SSRI studies [Blog post]. Retrieved from https://replicationindex.com/2015/08/05/using-the-r-index-to-detect-questionable-research-practices-in-ssri-studies/

Schimmack, U. (2016, January 31). The Replicability-Index: Quantifying statistical research integrity. Retrieved from https://replicationindex.com/2016/01/31/a-revised-introduction-to-the-r-index/

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9,* 666-681. doi: 10.1177/1745691614553988

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*, 203-220. doi: 10.3758/BF03196157

Strouse, G. A. (2019). Using digital media to support language learning in early childhood. In J. S. Horst & J. von Koss Torkildsen (Eds.), *International handbook of language acquisition* (pp. 485-503). New York, NY: Routledge.

Strouse, G. A., Newland, L. A., & Mourlam, D. J. (2019). Educational and fun? Parent versus preschooler perceptions and co-use of digital and print media. *AERA Open, 5.* doi: 10.1177/2332858419861085

Strouse, G. A., & Troseth, G. L. (2014). Supporting toddlers' transfer of word learning from video. *Cognitive Development*, *30*, 47-64. doi: 10.1016/j.cogdev.2014.01.002

Strouse, G. A., Troseth, G. L., O'Doherty, K. D., & Saylor, M. M. (2018). Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of Experimental Child Psychology*, *166*, 310-326. doi: 10.1016/j.jecp.2017.09.005

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science, 23,* 121-126. doi: 10.1177/0963721414522813

Tarasuik, J. C., Galligan, R., & Kaufman, J. (2011). Almost being there: Video communication with young children. *PLOS One, 6,* e17129. doi: 10.1371/journal.pone.0017129

Troseth, G. L. (2003a). Getting a clear picture: Young children's understanding of a televised image. *Developmental Science, 6,* 247-253. doi: 10.1111/1467-7687.00280

Troseth, G. L. (2003b). TV guide: Two-year-old children learn to use video as a source of information. *Developmental Psychology, 39*, 140-150. doi: 10.1037/0012-1649.39.1.140

Troseth, G. L. (2010). Is it life or is it Memorex? Video as a representation of reality. *Developmental Review*, *30*, 155-175. doi: 10.1016/j.dr.2010.03.007

Troseth, G. L., Casey, A. M., Lawver, K. A., Walker, J. M., & Cole, D. A. (2007). Naturalistic experience and the early use of symbolic artifacts. *Journal of Cognition and Development*, *8*, 309-331. doi: 10.1080/15248370701446772

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development*, *69*, 950–965. doi: 10.1111/j.1467-8624.1998.tb06153.x

Troseth, G. L., Flores, I., & Stuckelman, Z. D. (2019). When representation becomes reality: Interactive digital media and symbolic development. *Advances in Child Development and Behavior*, *56,* 65-108. doi: 10.1016/bs.acdb.2018.12.001

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development*, *77*, 786–799. doi: 10.1111/j.1467-8624.2006.00903.x

Troseth, G. L., Strouse, G. A., Verdine, B. N., & Saylor, M. M. (2018). Let's chat: On-screen social responsiveness is not sufficient to support toddlers' word learning from video. *Frontiers in Psychology, 9*. doi: 10.3389/fpsyg.2018.02195

Uhls, Y. T., Michikyan, M., Morris, J., Garcia, D., Small, G. W., Zgourou, E., & Greenfield, P. M. (2014). Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues. *Computers in Human Behavior*, *39*, 387-392. doi: 10.1016/j.chb.2014.05.036

Ulferts, H., Wolf, K. M., & Anders, Y. (2019). Impact of process quality in early childhood education and care on academic outcomes: Longitudinal meta-analysis. *Child Development, 90,* 1474-1489. doi: 10.1111/cdev.13296

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software, 36,* doi: 10.18637/jss.v036.i03

Zack, E., Barr, R., Gerhardstein, P., Dickerson, K., & Meltzoff, A. N. (2009). Infant imitation from television using novel touch screen technology. *British Journal of Development Psychology, 27,* 13–26. doi: 10.1111/sjop.12001

Zack, E., Barr, R., Gerhardstein, P., & Meltzoff, A. N. (2013). Fifteen-month-olds' transfer of learning between touch screen and real-world displays: Language cues and cognitive loads. *Scandinavian Journal of Psychology, 54*, 20-25. doi: 10.1111/sjop.12001

Zill, N. (2001). Does Sesame Street enhance school readiness?: Evidence from a national survey of children. In S. M. Fisch & R. T. Truglio (Eds.), *"G" is for growing* (pp. 115-130). Mahwah, NJ: Lawrence Erlbaum Associates.